Randomized Gossip Algorithm with Unreliable Communication[*]

Guodong Shi, Mikael Johansson and Karl Henrik Johansson[†]

Abstract

In this paper, we study an asynchronous randomized gossip algorithm under unreliable communication. At each instance, two nodes are selected to meet with a given probability. When nodes meet, two unreliable communication links are established with communication in each direction succeeding with a time-varying probability. It is shown that two particularly interesting cases arise when these communication processes are either perfectly dependent or independent. Necessary and sufficient conditions on the success probability sequence are proposed to ensure almost sure consensus or ϵ-consensus. Weak connectivity is required when the communication is perfectly dependent, while double connectivity is required when the communication is independent. Moreover, it is proven that with odd number of nodes, average preserving turns from almost forever (with probability one for all initial conditions) for perfectly dependent communication, to almost never (with probability zero for almost all initial conditions) for the independent case. This average preserving property does not hold true for general number of nodes. These results indicate the fundamental role the node interactions have in randomized gossip algorithms.

Keywords: Gossip algorithms, Unreliable communication, Consensus, Threshold

1 Introduction

Peer-to-peer networks, sensor networks and social networks constitute a new generation of engineering systems that has grown tremendously in importance over the last decade. The absence

---

[†]The authors are with ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology, Stockholm 10044, Sweden. Email: {`guodongs, mikaelj, kallej`}@kth.se

of a central decision-making entity, the unprecedented number of interacting nodes, the time-varying topology of node interactions, and the unreliability of nodes are key challenges for the analysis and design of these systems. Gossiping algorithms, in which each node exchanges data and decisions with at most one neighboring node in each time slot, have proven to be a robust and efficient way to structure distributed computation and information processing over such networks [10, 9, 11, 14]. A limitation of the current literature is that while it allows node interactions to be random, it assumes that the probabilities that two specific nodes interact is constant in time, and that when two nodes interact, both nodes completely and correctly execute the proposed algorithm. However, unreliable communication in wireless networks, and asymmetry of trust in social networks challenge such assumptions. This paper develops a framework for analysis of gossip algorithms that separates the random process for node interactions with the random process for successful information exchange and algorithm execution, and allows time-varying success probabilities for these operations. Necessary and sufficient conditions for the underlying graph structure and probability sequences are developed to ensure a.s. or ϵ-consensus under both perfectly dependent communication, in which the bidirectional message exchange between nodes either succeeds of fails, and independent communication, in which the success of communication in each direction is independent of the outcome in the other direction.

Gossip algorithms for distributed averaging arise in many applications, such as load balancing in parallel computing [18, 19], coordination of autonomous agents [22], distributed estimation [40] and analysis of opinion dynamics [43]. A central problem here is to analyze if a given gossip algorithm converges to consensus, and to determine the rate of convergence of the consensus process: Karp et al. [9] derived a general lower bound for synchronous gossiping; Kempe et al. [10] proposed a randomized gossiping algorithm on complete graphs and determined the order of its convergence rate; and Boyd et al. [12] established both lower and upper bounds for the convergence time of synchronous and asynchronous randomized gossiping algorithms, and developed algorithms for optimizing parameters to obtain fast consensus. In [43], a gossip algorithm was used to describe the spread of misinformation in social networks, where the state of each node was viewed as its belief and the randomized gossip algorithm characterized the dynamics of the belief evolution. A detailed introduction to gossip algorithms can be found in [14].

More generally, consensus problems on graphs have been investigated by researchers from many disciplines, including computer science [18, 19], engineering [27, 38, 20, 22, 21, 28] and

social sciences [17, 44]. Deterministic consensus algorithms have been studied extensively both for time-invariant and time-varying communication graphs. Efforts have typically been devoted to finding connectivity conditions which ensure convergence to consensus for the network model under consideration [21, 22, 42, 29, 20, 28, 25, 26]. Randomized consensus algorithms, motivated by the stochastic information flow over networks, have also been considered [30, 31, 34, 35, 32, 40]. Many sufficient and/or necessary conditions have been established that guarantee a global consensus with probability one [30, 31, 34, 35, 32, 39] under different settings for the randomization of the communication graph and convergence rates have also, in some cases, been established [34, 35, 36].

Most of the existing work on randomized consensus algorithms have focused on the influence of randomness in the underlying communication graph on the agreement seeking process. In [12, 33, 43], the communication graph was determined by a stochastic matrix whose elements specify the probability that each pair of nodes is selected to execute the gossip algorithm with each other in a given time instant. In [37], the communication graph was described by a finite-state Markov chain where each state corresponds to a specific realization of the communication graph, and conditions for guaranteeing almost sure consensus were given. In [30], the authors studied linear consensus dynamics with communications graphs given by a sequence of independent and identically distributed (i.i.d.) Erdös-Rényi random graphs; the analysis was later extended to directed Erdös-Rényi graphs in [31]. Mean-square performance for consensus algorithms over i.i.d. random graphs was studied in [34] and the influence of random packet drops was investigated in [34, 36]. Distributed consensus over random graphs with graph independence was also studied in [32, 40]. In all these works, the randomness in the underlying communication graph describes the randomness in the process of initiating interactions. Even if two nodes meet with each other, the information exchange could still fail, or be only partially executed (e.g., only by one of the two nodes) due to unreliable communication between nodes and the asymmetry of trust.

This paper studies an asynchronous randomized gossip algorithm under unreliable communication. A key feature of the proposed analysis framework is that it separates the random process for *initiating interactions* between node pairs from the random process for *successful communication* between two interacting nodes, and that it considers time-varying success probabilities for inter-node communication. Both perfectly dependent communication, in which the bi-directional message exchange between two nodes either succeeds or fails, and independent

communication, in which the success of communication in each direction is independent of the outcome in the other direction, are considered. The perfectly dependent and independent node communication reflects directly the symmetry of the gossiping algorithms [33]. We derive necessary and sufficient conditions on the underlying graph structure and the communication success probability sequence to ensure a.s. consensus or ϵ-consensus under perfectly dependent and independent communication, respectively. We investigate the impact of communication dependence on the behavior of the algorithm, and show how the underlying requirement for reaching a.s. consensus goes from weak connectivity in the perfectly dependent case to double connectivity in the independent case, while the probability of preserving the average decreases from one for perfectly dependent communication to zero for the independent case when the number of nodes is odd.

The rest of the paper is organized as follows. In Section 2, some preliminary concepts are introduced. We present the network model, the randomized gossip algorithm, and the standing assumptions in Section 3. Then main results for perfectly dependent and independent communication are given in Sections 4 and 5, respectively. Finally, concluding remarks are given in Section 6.

2 Preliminaries

In this section, we recall some basic definitions from graph theory [3], stochastic matrices [2] and Bernoulli trials [1].

2.1 Directed Graphs

A directed graph (digraph) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a finite set $\mathcal{V}$ of nodes and an arc set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. An element $e = (i, j) \in \mathcal{E}$ is called an *arc* from node $i \in \mathcal{V}$ to $j \in \mathcal{V}$. If the arcs are pairwise distinct in an alternating sequence $v_0 e_1 v_1 e_2 v_2 \ldots e_k v_k$ of nodes $v_i \in \mathcal{V}$ and arcs $e_i = (v_{i-1}, v_i) \in \mathcal{E}$ for $i = 1, 2, \ldots, k$, the sequence is called a (directed) *path* with *length* k. A path with no repeated nodes is called a *simple path*. A path from i to j is denoted as $i \to j$, and the length of $i \to j$ is denoted as $|i \to j|$. If there exists a path from node i to node j, then node j is said to be reachable from node i. Each node is thought to be reachable by itself. A node v from which any other node is reachable is called a *center* (or a *root*) of $\mathcal{G}$. A digraph $\mathcal{G}$ is said to be *strongly connected* if it contains path $i \to j$ and $j \to i$ for every pair of nodes i and j, and *quasi-strongly*

connected if $\mathcal{G}$ has a center [4].

The *converse graph*, $\mathcal{G}^T$ of a digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, is defined as the graph obtained by reversing the orientation of all arcs in $\mathcal{E}$. The *distance* from i to j in a digraph $\mathcal{G}$, $d(i, j)$, is the length of a shortest simple path $i \to j$ if j is reachable from i, and the *diameter* of $\mathcal{G}$ is diam$(\mathcal{G})$= $\max\{d(i,j)|i, j \in \mathcal{V}, \ j$ is reachable from $i\}$.

The union of two digraphs with the same node set $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_2)$ is defined as $\mathcal{G}_1 \cup \mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_1 \cup \mathcal{E}_2)$; we denote $\mathcal{G}_1 \subseteq \mathcal{G}_2$ if $\mathcal{E}_1 \subseteq \mathcal{E}_2$. A digraph $\mathcal{G}$ is said to be bidirectional if for every two nodes i and j, $(i, j) \in \mathcal{E}$ if and only if $(j, i) \in \mathcal{E}$. A bidirectional graph $\mathcal{G}$ is said to be *connected* if there is a path between any two nodes. A digraph $\mathcal{G}$ is said to be *weakly connected* if it is connected as a bidirectional graph when all the arc directions are ignored. Strongly or quasi-strongly connected digraphs are hence always weakly connected.

2.2   Stochastic Matrices

A finite square matrix $M = [m_{ij}] \in \mathbb{R}^{n \times n}$ is called *stochastic* if $m_{ij} \geq 0$ for all i, j and $\sum_j m_{ij} = 1$ for all i. For a stochastic matrix M, introduce

$$\delta(M) = \max_j \max_{\alpha, \beta} |m_{\alpha j} - m_{\beta j}|, \quad \lambda(M) = 1 - \min_{\alpha, \beta} \sum_j \min\{m_{\alpha j}, m_{\beta j}\}. \tag{1}$$

If $\lambda(M) < 1$ we call M a *scrambling* matrix. The following lemma can be found in [16].

Lemma 1 *For any $k \geq 1$ stochastic matrices $M_1, \ldots, M_k$,*

$$\delta(M_k \cdots M_2 M_1) \leq \prod_{i=1}^k \lambda(M_i). \tag{2}$$

A stochastic matrix $M = [m_{ij}] \in \mathbb{R}^{n \times n}$ is called *doubly stochastic* if also M^T is stochastic.

Let $P = [p_{ij}] \in \mathbb{R}^{n \times n}$ be a matrix with nonnegative entries. We can associate a unique digraph $\mathcal{G}_P = \{\mathcal{V}, \mathcal{E}_P\}$ with P on node set $\mathcal{V} = \{1, \ldots, n\}$ such that $(j, i) \in \mathcal{E}_P$ if and only if $p_{ij} > 0$. We call $\mathcal{G}_P$ the *induced graph* of P.

2.3   Bernoulli Trials

A sequence of independently distributed Bernoulli trials is a finite or infinite sequence of independent random variables $\mathfrak{B}_0, \mathfrak{B}_1, \mathfrak{B}_2, \ldots$, such that

(i) For each $k \geq 0$, $\mathfrak{B}_k$ equals either 0 or 1;

(ii) For each $k \geq 0$, the probability that $\mathfrak{B}_k = 1$ is p_k.

We call p_k the success probability for time k. The sequence of integers

$$0 \leq \zeta_1 < \zeta_2 < \cdots : \quad \mathfrak{B}_{\zeta_m} = 1 \tag{3}$$

is called the *Bernoulli (success)* sequence associated with the sequence of Bernoulli trials with ζ_m marking the time of the m'th success.

3 Problem Definition

In this section, we present the considered network model and define the problem of interest.

3.1 Node Pair Selection Process

Consider a network with node set $\mathcal{V} = \{1, \ldots, n\}$ ($n \geq 3$). Let the digraph $\mathcal{G}_0 = (\mathcal{V}, \mathcal{E}_0)$ denote the *underlying* graph of the considered network. The underlying graph indicates potential interactions between nodes. We use the asynchronous time model introduced in [12] to describe node interactions. Each node meets other nodes at independent time instances defined by a rate-one Poisson process. This is to say, the inter-meeting times at each node follows a rate-one exponential distribution. Without loss of generality, we can assume that at most one node is active at any given instance. Let $x_i(k) \in \mathbb{R}$ denote the state (value) of node i at the k'th meeting slot among all the nodes.

Node interactions are characterized by an $n \times n$ matrix $A = [a_{ij}]$, where $a_{ij} \geq 0$ for all $i, j = 1, \ldots, n$ and $a_{ij} > 0$ if and only if $(j, i) \in \mathcal{E}_0$. We assume A is a stochastic matrix. The meeting process is defined as follows.

The node pair selection process for the gossip algorithm is defined as follows.

Definition 1 *(Node Pair Selection Process) At each time $k \geq 0$,*

(i) A node $i \in \mathcal{V}$ is drawn with probability $1/n$;

(ii) Node i picks the pair (i, j) with probability a_{ij}.

Note that, by the definition of the node pair selection process, the underlying graph $\mathcal{G}_0$ is actually the same as $\mathcal{G}_A$, the induced graph of the node pair selection matrix A. For $\mathcal{G}_0$, we use the following assumption.

6

A1. *(Weak Connectivity)* The underlying graph $\mathcal{G}_0$ is weakly connected.

Remark 1 *Node that, node pairs (i, j) and (j, i) have different meaning according to the node pair selection process. When (i, j) is selected, node i is the node first picked and then i picks j. While pair (j, i) is selected means j is first picked who picks i later.*

Remark 2 *The node pair selection matrix A being a stochastic matrix has natural meaning that node i's decisions form a well-defined probability space when it is selected at time k. However, it is not essential for the following discussions in the sense that all the results still stand even this assumption is replaced by the condition that each row sum of A is no larger than one. The same assumption is made in [12, 43].*

Remark 3 *Let $\mathcal{G}_{A+A^T}$ denote the induced graph of matrix $A + A^T$. Apparently $A + A^T$ is symmetric, and thus $\mathcal{G}_{A+A^T}$ is a bidirectional graph. It is not hard to see that $\mathcal{G}_0$ is weakly connected if and only if $\mathcal{G}_{A+A^T}$ is a connected bidirectional graph.*

Remark 4 *In the standing assumption of [12], the matrix A is supposed to have its largest eigenvalue equal to 1 and all other $n - 1$ eigenvalues strictly less than 1 in magnitude. This condition is equivalent with that $\mathcal{G}_0$ is quasi-strongly connected [28, 25]. On the other hand, in [43], $\mathcal{G}_0$ is assumed to be strongly connected. Therefore, Assumption 1 is a weaker assumption, compared to the one in [12, 43].*

Remark 5 *In order to guarantee convergence for the gossip algorithm discussed below, A1 cannot be further weakened based on the following argument. Let us just assume A1 does not hold true. Then there will be two disjoint node sets $\mathcal{V}_1, \mathcal{V}_2 \subset \mathcal{V}$ such that there is no link connecting the two sets. As a result, nodes in $\mathcal{V}_i, i = 1, 2$ can only communicate with nodes belonging to the same subset. Therefore, the network is essentially divided into two isolated parts, and a convergence for the whole network is thus impossible.*

3.2 Node Communication Process

When pair (i, j) is selected, both nodes try to set their states equal to the average of their current states. To this end, two communication links with opposite directions are established between the two nodes.

Let $\{P_k^+\}_0^\infty$ and $\{P_k^-\}_0^\infty$ be two deterministic sequences with $0 \le P_k^+, P_k^- \le 1$ for all k. The node communication process is defined as follows.

Definition 2 *(Node Communication Process) Independent with time, node states, and the node pair selection process,*

(i) $\mathbf{P}\big(\mathbb{E}_k^+\big) = P_k^+$ *with* $\mathbb{E}_k^+ = \big\{node\ i\ receives\ x_j(k)\ when\ (i,j)\ is\ selected\ at\ time\ k\big\}$;

(ii) $\mathbf{P}\big(\mathbb{E}_k^-\big) = P_k^-$ *with* $\mathbb{E}_k^- = \big\{node\ j\ receives\ x_i(k)\ when\ (i,j)\ is\ selected\ at\ time\ k\big\}$.

If a node fails to receive the value of the other node, it will keep its current state. Note that we do not in general impose independence between i receiving $x_j(k)$ and j receiving $x_i(k)$ when pair (i,j) is selected. In fact, we will study how such potential dependence in the communication process influences the convergence of the gossip algorithm.

Remark 6 *A randomized gossip algorithm can also be viewed as belief propagation in a social network, where $x_i(k)$ represents the belief of node i. Then the communication process naturally captures the loss of 'trust' when two nodes meet and exchange opinions [43, 44]. Therefore, from a social network viewpoint, the discussion in this paper on the convergence property of the gossip algorithm establishes the influence of missing 'trust' in belief agreement.*

3.3 Problem

Let the initial condition be $x^0 = x(k_0) = (x_1(k_0) \dots x_n(k_0))^T \in \mathbb{R}^n$, where $k_0 \geq 0$ is an arbitrary integer. According to the node pair selection process and the node communication process, the iteration of the gossip algorithm can be expressed as: for $k \geq k_0$,

$$x_i(k+1) = \begin{cases} \frac{1}{2}x_i(k) + \frac{1}{2}x_j(k), & \text{if } \mathcal{M}^{\langle i,j \rangle} \text{ happens} \\ x_i(k), & \text{otherwise,} \end{cases} \tag{4}$$

where

$$\mathcal{M}_k^{\langle i,j \rangle} \doteq \Big\{(i,j) \text{ is selected or } (j,i) \text{ is selected, and } i \text{ receives } x_j(k) \text{ at time } k\Big\}$$

denotes the event that node i successfully updates at time k. According to the definitions above, we have

$$\mathbf{P}\Big(\mathcal{M}_k^{\langle i,j \rangle}\Big) = \frac{a_{ij}}{n}P_k^+ + \frac{a_{ji}}{n}P_k^-; \qquad \mathbf{P}\Big(\mathcal{M}_k^{\langle j,i \rangle}\Big) = \frac{a_{ji}}{n}P_k^+ + \frac{a_{ij}}{n}P_k^-.$$

Therefore, the two events, $\mathcal{M}_k^{\langle i,j \rangle}$ and $\mathcal{M}_k^{\langle j,i \rangle}$, are not necessarily symmetric in their probabilities, due to the potential asymmetry of the meeting matrix A.

In this paper, we study the convergence of the randomized gossip consensus algorithm and the time it takes for the network to reach a consensus. Let

$$x(k; k_0, x^0) = \Big(x_1\big(k; k_0, x_1(k_0)\big) \dots x_n\big(k; k_0, x_n(k_0)\big) \Big)^T \in \mathbb{R}^n$$

be the random process driven by the randomized algorithm (4). When it is clear from the context, we will identify $x(k; k_0, x^0)$ with $x(k)$.

Denote

$$H(k) \doteq \max_{i=1,\dots,n} x_i(k), \quad h(k) \doteq \min_{i=1,\dots,n} x_i(k)$$

as the maximum and minimum states among all nodes, respectively, and define $\mathcal{H}(k) \doteq H(k) - h(k)$ as the consensus metric. We introduce the following definition.

Definition 3 *(i) A global a.s. consensus is achieved if*

$$\mathbf{P}\big(\lim_{k \to \infty} \mathcal{H}(k) = 0\big) = 1 \tag{5}$$

for any initial condition $x^0 \in \mathbb{R}^n$.

(ii) Let the ϵ-computation time be

$$T_{\mathrm{com}}(\epsilon) \doteq \sup_{x(k_0)} \inf \Big\{ k - k_0 : \quad \mathbf{P}\Big(\frac{\mathcal{H}(k)}{\mathcal{H}(k_0)} \geq \epsilon \Big) \leq \epsilon \Big\}. \tag{6}$$

Then a global a.s. ϵ-consensus is achieved if

$$T_{\mathrm{com}}(\epsilon) = O(\log \epsilon^{-1}) \tag{7}$$

where by definition $f(\epsilon) = O\big(g(\epsilon)\big)$ means that $\limsup_{\epsilon \to 0} f(\epsilon)/g(\epsilon) < \infty$ is a nonzero constant.

Remark 7 *A global a.s. only requires that $\mathcal{H}(t)$ will converge to zero with probability one. If it is further required that the convergence speed is sufficiently fast, we use global a.s. ϵ-agreement. The ϵ-computation $T_{\mathrm{com}}(\epsilon)$ is essentially equivalent with the definition of ϵ-averaging time in [12], which is a probabilistic version of similar concepts used to characterize the convergence rate of deterministic consensus algorithms in the literature, e.g., [42].*

Recall that until now, when i is selected to meet node j at time k, no assumption has been made on the dependence between the communication from i to j, and the one from j to i. In the following two sections, we will discuss the convergence of the considered randomized gossip algorithm with perfectly dependent and independent communication, respectively. We will show that the dependence in the node communication plays a critical role in determining the behavior of the gossip algorithm.

4 Convergence under Perfectly Dependent Communication

In this section, we study the case when the communication between nodes i and j is perfectly dependent, as described in the following assumption.

A2. *(Perfectly Dependent Communication)* The events $\mathbb{E}_k^+ = \mathbb{E}_k^-$ except for a set with probability zero for all k.

Note that A2 is equivalent to assuming that $\mathbf{P}(\mathbb{E}_k^+|\mathbb{E}_k^-) = \mathbf{P}(\mathbb{E}_k^-|\mathbb{E}_k^+) = 1$. Hence, we have $P_k^+ = P_k^-$ and at each instant, with probability $P_k \doteq P_k^+ = P_k^-$, both $\mathbb{E}_k^+$ and $\mathbb{E}_k^-$ occur, and with probability $1 - P_k$ they both fail. With A2, the gossip algorithm can be expressed as

$$x(k+1) = W(k)x(k), \tag{8}$$

where $W(k)$ is the random matrix satisfying

$$\mathbf{P}\Big(W(k) = W_{\langle ij \rangle} \doteq I - \frac{(e_i - e_j)(e_i - e_j)^T}{2}\Big) = \frac{a_{ij} + a_{ji}}{n} P_k, \quad i \neq j \tag{9}$$

with $e_m = (0\ldots 0\,1\,0\ldots 0)^T$ denoting the $n \times 1$ unit vector whose m'th component is 1. Moreover, $\mathbf{P}\big(W(k) = W_{\langle ii \rangle} = I\big) = 1 - \sum_{i>j} \frac{a_{ij}+a_{ji}}{n} P_k$.

The main result on a.s. consensus for the considered gossip algorithm under perfectly dependent communication is stated as follows.

Theorem 1 *Suppose A1 (Weak Connectivity) and A2 (Perfectly Dependent Communication) hold. Global a.s. consensus is achieved if and only if $\sum_{k=0}^{\infty} P_k = \infty$.*

Denote $D = \text{diag}(d_1 \ldots d_n)$ with $d_i = \sum_{j=1}^{n}(a_{ij} + a_{ji})$. For a.s. ϵ-consensus, we have the following conclusion.

Theorem 2 *Suppose A1 (Weak Connectivity) and A2 (Perfectly Dependent Communication) hold. Global a.s. ϵ-consensus is achieved if and only if there exist a constant $p_* > 0$ and an integer $T_* \geq 1$ such that $\sum_{k=m}^{m+T_*-1} P_k \geq p_*$ for all $m \geq 0$. In fact, we have*

$$T_{\text{com}}(\epsilon) \leq 3\Big[\log\big(1 - \frac{\lambda_2^* p_*}{2nT_*}\big)^{-1}\Big]^{-1} \log \epsilon^{-1} + O(1), \tag{10}$$

where λ_2^* is the second smallest eigenvalue of $D - (A + A^T)$.

Theorem 1 indicates that $\sum_{k=0}^{\infty} P_k = \infty$ is actually a threshold for the gossip algorithm to reach a.s. consensus. For ϵ-consensus, Theorem 2 implies that $\sum_{k=0}^{m} P_k = O(m)$ is the threshold condition, which actually requires that $\sum_{k=0}^{m} P_k$ grows linearly as a function of m.

Remark 8 *Theorems 1 and 2 rely on the fact that there are at least three nodes in the network. If the network contains only two nodes, then both Theorems 1 and 2 no longer hold true. This phenomenon is interesting since many consensus results in the literature are independent of the number of nodes, e.g., [22, 28, 21, 25, 26, 27].*

Let the random variable $\xi(k_0, x^0)$ denote the consensus limit (supposed to exist), i.e.,

$$\lim_{k \to \infty} x_i(k) = \xi, \quad a.s. \quad i = 1, \ldots, n. \tag{11}$$

Denote $x_{\text{ave}} = \sum_{i=1}^{n} x_i(k_0)/n$ be the average of the initial values. Then the following conclusion holds showing that the average is preserved almost surely with perfectly dependent communication.

Theorem 3 *Suppose A1 (Weak Connectivity) and A2 (Perfectly Dependent Communication) hold. Then for all initial conditions $x^0 = x(k_0) \in \mathbb{R}^n$, we have*

$$\mathbf{P}\Big(\sum_{i=1}^{n} x_i(k) = n x_{\text{ave}}, \quad k \geq k_0 \Big) = 1. \tag{12}$$

Consequently, we have $\mathbf{P}\big(\xi = x_{\text{ave}}\big) = 1$ whenever the consensus limit exists.

In the following two subsections, we will present the proof of Theorems 1 and 2, respectively. Theorem 3 follows from the proof Theorem 1.

The upcoming analysis relies on the following well-known lemmas.

Lemma 2 *Suppose $0 \leq b_k < 1$ for all k. Then $\sum_{k=0}^{\infty} b_k = \infty$ if and only if $\prod_{k=0}^{\infty}(1 - b_k) = 0$.*

Lemma 3 $\log(1 - t) \geq -2t$ *for all $0 \leq t \leq 1/2$.*

4.1 Proof of Theorem 1

(Sufficiency.) This part of the proof is based on a similar argument as is used in [12]. Define $L(k) = \sum_{i=1}^{n} |x_i(k) - x_{\text{ave}}|^2$, where $|\cdot|$ represents the Euclidean norm of a vector or the absolute value of a scalar.

It is easy to verify for every possible sample and fixed instant k that $W_{\langle ij \rangle}$ of the random matrix $W(k)$ defined in (8) and (9) fulfills

(i). $W_{\langle ij \rangle}$ is a doubly stochastic matrix, i.e., $W_{\langle ij \rangle}\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T W_{\langle ij \rangle} = \mathbf{1}^T$;

(ii). $W_{\langle ij \rangle}$ is a projection matrix, i.e., $W_{\langle ij \rangle} = W_{\langle ij \rangle}^T W_{\langle ij \rangle}$.

Therefore, we have

$$
\begin{aligned}
\mathbf{E}\Big(L(k+1)\big|x(k)\Big) &= \mathbf{E}\Big(\big(x(k+1)-x_{\mathrm{ave}}\mathbf{1}\big)^T\big(x(k+1)-x_{\mathrm{ave}}\mathbf{1}\big)\big|x(k)\Big) \\
&= \mathbf{E}\Big(\big(W(k)x(k)-x_{\mathrm{ave}}\mathbf{1}\big)^T\big(W(k)x(k)-x_{\mathrm{ave}}\mathbf{1}\big)\big|x(k)\Big) \\
&= \mathbf{E}\Big(\big(x(k)-x_{\mathrm{ave}}\mathbf{1}\big)^T W(k)^T W(k)\big(x(k)-x_{\mathrm{ave}}\mathbf{1}\big)\big|x(k)\Big) \\
&= \big(x(k)-x_{\mathrm{ave}}\mathbf{1}\big)^T \mathbf{E}\big(W(k)^T W(k)\big)\big(x(k)-x_{\mathrm{ave}}\mathbf{1}\big) \\
&= \big(x(k)-x_{\mathrm{ave}}\mathbf{1}\big)^T \mathbf{E}\big(W(k)\big)\big(x(k)-x_{\mathrm{ave}}\mathbf{1}\big)
\end{aligned}
\tag{13}
$$

Since $W(k)$ is doubly stochastic, we know that the sum of the nodes' states, $\sum_{i=1}^{n} x_i(k)$, is preserved with probability one, and $\mathbf{1}$ is the eigenvector corresponding to eigenvalue 1 of $\mathbf{E}\big(W(k)\big)$ (Theorem 3 therefore holds). Thus, we can conclude from (13) that

$$
\mathbf{E}\Big(L(k+1)\big|x(k)\Big) \le \lambda_2\Big(\mathbf{E}\big(W(k)\big)\Big)\big(x(k)-x_{\mathrm{ave}}\mathbf{1}\big)^T\big(x(k)-x_{\mathrm{ave}}\mathbf{1}\big) = \lambda_2\Big(\mathbf{E}\big(W(k)\big)\Big)L(k), \tag{14}
$$

where $\lambda_2(M)$ for a stochastic matrix M denotes the largest eigenvalue in magnitude except for the eigenvalue at one. Here note that $\mathbf{E}\big(W(k)\big)$ is symmetric and positive semi-definite.

Now according to (9), we see that

$$
\mathbf{E}\big(W(k)\big) = I - \frac{P_k}{2n}\big(D-(A+A^T)\big). \tag{15}
$$

Note that $D-(A+A^T)$ is actually the (weighted) Laplacian of the graph $\mathcal{G}_{A+A^T}$. With assumption A1, $\mathcal{G}_{A+A^T}$ is a connected graph (cf., Remark 3), and therefore, based on the well-known property of Laplacian matrix of connected graphs [3], we have $\lambda_2^* > 0$, where λ_2^* is the second smallest eigenvalue of $D-(A+A^T)$. On the other hand, since A is a stochastic matrix, it is straightforward to see that

$$
\sum_{j=1,j\neq i} a_{ij}+a_{ji} \le n \tag{16}
$$

for all $i=1,\dots,n$. According to Gershgorin circle theorem, all the eigenvalues of $D-(A+A^T)$ are bounded by $2n$. Therefore, we conclude from (15) that for all k,

$$
\lambda_2\Big(\mathbf{E}\big(W(k)\big)\Big) = 1 - \frac{\lambda_2^*}{2n}P_k. \tag{17}
$$

With (14) and (17), we obtain

$$
\mathbf{E}\Big(L(k+1)\Big) \le \prod_{i=k_0}^{k} \lambda_2\Big(\mathbf{E}\big(W(i)\big)\Big)L(k_0) = \prod_{i=k_0}^{k}\Big(1-\frac{\lambda_2^*}{2n}P_i\Big)L(k_0), \tag{18}
$$

Therefore, based on Lemma 2 and Fatou's lemma, we have

$$\mathbf{E}\Big(\lim_{k\to\infty} L(k)\Big) \leq \lim_{k\to\infty} \mathbf{E}\Big(L(k)\Big) = 0, \tag{19}$$

where $\lim_{k\to\infty} L(k)$ exists simply from the fact that the sequence is non-increasing. This immediately implies

$$\mathbf{P}\Big(\lim_{k\to\infty} x_i(k) = x_{\text{ave}}\Big) = 1. \tag{20}$$

The sufficiency claim of the theorem thus holds.

(Necessity.) From the definition of the gossip algorithm, we have

$$\mathbf{P}\big(x_i(k+1) = x_i(k)\big) \geq 1 - \mathbf{P}\big(i \text{ receives } x_j(k)\big) \cdot \sum_{j=1,\ j\neq i}^{n} \Big[\mathbf{P}\big(\text{pair } (i,j) \text{ is selected}\big)$$

$$+ \mathbf{P}\big(\text{pair } (j,i) \text{ is selected}\big)\Big]$$

$$= 1 - P_k \sum_{j=1,\ j\neq i}^{n} \frac{1}{n}\big(a_{ij} + a_{ji}\big)$$

$$\doteq 1 - h_i P_k, \tag{21}$$

where $h_i = \sum_{j=1,\ j\neq i} \frac{1}{n}\big(a_{ij} + a_{ji}\big), i = 1, \ldots, n$. Noting the fact that

$$\sum_{i=1}^{n} h_i = \sum_{i=1}^{n} \sum_{j=1,\ j\neq i}^{n} \frac{1}{n}\big(a_{ij} + a_{ji}\big) = 2 - \sum_{i=1}^{n} a_{ii} \leq 2, \tag{22}$$

there exists at least one node $\alpha_1 \in \mathcal{V}$ such that $h_{\alpha_1} < 1$ since $n \geq 3$. Moreover, assumption A1 further guarantees that $h_i > 0, i = 1, \ldots, n$, which implies that there exists another node $\alpha_2 \in \mathcal{V}$ such that $h_{\alpha_2} < 1$.

Therefore, if $\sum_{k=0}^{\infty} P_k < \infty$, we have

$$\mathbf{P}\big(x_{\alpha_i}(k) = x_{\alpha_i}(k_0), k \geq k_0\big) = \prod_{k=k_0}^{\infty} \big(1 - h_{\alpha_i} P_k\big) \doteq \sigma_i > 0, \quad i = 1, 2 \tag{23}$$

based on Lemma 2. Consequently, choosing $x_{\alpha_1}(k_0) \neq x_{\alpha_2}(k_0)$, consensus will fail with probability $\sigma_1 \sigma_2 > 0$. This completes the proof.

4.2 Proof of Theorem 2

(Sufficiency.) Recall that $\mathcal{H}(k) \doteq \max_{i=1,\ldots,n} x_i(k) - \min_{i=1,\ldots,n} x_i(k)$. Suppose nodes α and β reach the maximum and minimum values at time k, respectively, i.e.,

$$x_\alpha(k) = \max_{i=1,\ldots,n} x_i(k); \quad x_\beta(k) = \min_{i=1,\ldots,n} x_i(k).$$

Then we have

$$L(k) = \sum_{i=1}^{n} |x_i(k) - x_{\text{ave}}|^2 \geq |x_\alpha(k) - x_{\text{ave}}|^2 + |x_\beta(k) - x_{\text{ave}}|^2$$

$$\geq \frac{1}{2}|x_\alpha(k) - x_\beta(k)|^2$$

$$= \frac{1}{2}\mathcal{H}^2(k). \tag{24}$$

On the other hand,

$$L(k_0) = \sum_{i=1}^{n} |x_i(k_0) - x_{\text{ave}}|^2 \leq \frac{1}{n^2}\Big(\sum_{i=1}^{n}\Big|nx_i(k_0) - \sum_{j=1}^{n} x_j(k_0)\Big|\Big)^2$$

$$\leq \frac{1}{n^2}\Big(\sum_{i=1}^{n}\sum_{j=1,\ j\neq i}^{n}\Big|x_i(k_0) - x_j(k_0)\Big|\Big)^2$$

$$\leq \frac{n-1}{n}\mathcal{H}^2(k_0). \tag{25}$$

With (24) and (25) and applying Markov's inequality, we have

$$\mathbf{P}\Big(\frac{\mathcal{H}(k)}{\mathcal{H}(k_0)} \geq \epsilon\Big) = \mathbf{P}\Big(\frac{\mathcal{H}^2(k)}{\mathcal{H}^2(k_0)} \geq \epsilon^2\Big)$$

$$\leq \mathbf{P}\Big(\frac{L(k)}{L(k_0)} \geq \frac{n}{2(n-1)}\epsilon^2\Big)$$

$$\leq \frac{2(n-1)}{n}\epsilon^{-2}\frac{\mathbf{E}(L(k))}{L(k_0)}$$

$$\leq \frac{2(n-1)}{n}\epsilon^{-2}\prod_{i=k_0}^{k-1}\Big(1 - \frac{\lambda_2^*}{2n}P_i\Big) \tag{26}$$

where the last inequality holds from (18). Since $\sum_{k=m}^{m+T_*-1} P_k \geq p_*$ for all $m \geq 0$, according to the arithmetic mean-geometric mean inequality, we have that for all $m \geq 0$,

$$\prod_{k=m}^{m+T_*-1}\Big(1 - \frac{\lambda_2^*}{2n}P_k\Big) \leq \Big(\frac{T_* - \frac{\lambda_2^*}{2n}\sum_{k=m}^{m+T_*-1} P_k}{T_*}\Big)^{T_*} \leq \Big(1 - \frac{\lambda_2^* p_*}{2nT_*}\Big)^{T_*} \doteq c_* < 1. \tag{27}$$

As a result, we obtain

$$\prod_{i=k_0}^{k-1}\Big(1 - \frac{\lambda_2^*}{2n}P_i\Big) \leq c_*^{\lfloor\frac{k-k_0}{T_*}\rfloor} \leq c_*^{\frac{k-k_0}{T_*}-1}, \tag{28}$$

where $\lfloor z \rfloor$ denotes the largest integer no larger than z.

Then (26) and (28) lead to

$$\mathbf{P}\Big(\frac{\mathcal{H}(k)}{\mathcal{H}(k_0)} \geq \epsilon\Big) \leq \frac{2(n-1)}{n}\epsilon^{-2}c_*^{\frac{k-k_0}{T_*}-1}, \tag{29}$$

14

which implies

$$T_{\text{com}}(\epsilon) \leq T_* \left[\frac{3 \log \epsilon^{-1} + \log \frac{2(n-1)}{c_* n}}{\log c_*^{-1}} \right] = 3 \left[\log \left(1 - \frac{\lambda_2^* p_*}{2n T_*} \right)^{-1} \right]^{-1} \log \epsilon^{-1} + O(1). \qquad (30)$$

The desired conclusion follows.

(Necessity.) We prove the necessity part of Theorem 2 by a contradiction argument. Let α_1, α_2 be defined as in the proof of Theorem 1. Set $x_{\alpha_1}(k_0) = 0$, $x_{\alpha_2}(k_0) = 1$ and $x_j(k_0) \in [0,1]$ for all other nodes. Then according to (23), we have

$$\mathbf{P}\Big(\frac{\mathcal{H}(k)}{\mathcal{H}(k_0)} \geq \epsilon \Big) \geq \mathbf{P}\big(x_{\alpha_i}(t) = x_{\alpha_i}(k_0), \ i = 1,2; k_0 \leq t \leq k \big) = \prod_{t=k_0}^{k-1} \big(1 - h_{\alpha_1} P_t \big)\big(1 - h_{\alpha_2} P_t \big)$$

$$(31)$$

Take $\epsilon = 1/\ell$ with $\ell = 1, 2, \dots$. Suppose suitable p_* and T_* cannot be found such that $\sum_{k=m}^{m+T_*-1} P_k \geq p_*$ for all $m \geq 0$. Then for any $\hat{T} = \ell \log \ell$, there exists an integer $\hat{k} \geq 0$ such that $\sum_{t=\hat{k}}^{\hat{k}+\hat{T}} P_t < 1/2$. According to (31) and Lemma 3, we have

$$\begin{aligned}
\mathbf{P}\Big(\frac{\mathcal{H}(\hat{k} + \hat{T} + 1)}{\mathcal{H}(\hat{k})} \geq \epsilon \Big) &\geq \prod_{t=\hat{k}}^{\hat{k}+\hat{T}} \big(1 - h_{\alpha_1} P_t \big)\big(1 - h_{\alpha_2} P_t \big) \\
&= e^{\sum_{t=\hat{k}}^{\hat{k}+\hat{T}} \big(\log(1 - h_{\alpha_1} P_t) + \log(1 - h_{\alpha_2} P_t) \big)} \\
&\geq e^{-2(h_{\alpha_1} + h_{\alpha_2}) \sum_{t=\hat{k}}^{\hat{k}+\hat{T}} P_t} \\
&> e^{-(h_{\alpha_1} + h_{\alpha_2})} \\
&\geq 1/\ell \qquad\qquad\qquad\qquad\qquad (32)
\end{aligned}$$

for all $\ell \geq e^{h_{\alpha_1} + h_{\alpha_2}}$. This immediately implies $T_{\text{com}}(1/\ell) \geq \ell \log \ell$, which suggests that $T_{\text{com}}(\epsilon) = O(\log \epsilon^{-1})$ does not hold. The proof has been completed.

5 Convergence under Independent Communication

In this section, we focus on the case when the communication between nodes i and j is independent, as described in the following assumption.

A3. *(Independent Communication)* The events $\mathbb{E}_k^+$ and $\mathbb{E}_k^-$ are independent for all k.

Remark 9 *Symmetric and asymmetric randomized gossip algorithms were studied in [33]. The symmetry in [33] is a deterministic concept where the gossip algorithm is either symmetric or*

15

asymmetric, and therefore, it is binary. While the dependence discussed in this paper carefully characterizes how much symmetry is missing from a probabilistic viewpoint. The symmetric model in [33] is a special case of ours when $P_k^+ = P_k^- = 1$ for all k (in this case perfectly dependent communication coincides with independent communication). The asymmetric model in [33] is a special case of our independent communication model with $P_k^+ = 1$ and $P_k^- = 0$ for all k.

With A3, $\mathbb{E}_0^+, \mathbb{E}_0^-, \mathbb{E}_1^+, \mathbb{E}_1^-, \ldots$ is a sequence of independent events, and the considered gossip algorithm can be expressed as

$$x(k+1) = W(k)x(k), \tag{33}$$

where $W(k)$ is the random matrix satisfying

(i) $\mathbf{P}\left(W(k) = I - \frac{e_i(e_i - e_j)^T}{2}\right) = \frac{a_{ij}}{n}P_k^+(1 - P_k^-) + \frac{a_{ji}}{n}P_k^-(1 - P_k^+), \quad i \neq j;$

(ii) $\mathbf{P}\left(W(k) = I - \frac{(e_i - e_j)(e_i - e_j)^T}{2}\right) = \frac{a_{ij} + a_{ji}}{n}P_k^+ P_k^-, \quad i > j;$

(iii) $\mathbf{P}\left(W(k) = I\right) = 1 - \sum_{i>j}\frac{a_{ij} + a_{ji}}{n}\left(P_k^+ + P_k^- - P_k^+ P_k^-\right).$

In order to establish the convergence results under independent communication, we need the following condition for the underlying connectivity.

A4. *(Double Connectivity)* Both the underlying graph $\mathcal{G}_0$ and its converse graph $\mathcal{G}_0^T$ are quasi-strongly connected.

Remark 10 *Note that the condition of $\mathcal{G}_0$ being strongly connected implies A4, but not vice versa. Moreover, it is not hard to see that $\mathcal{G}_0 = \mathcal{G}_A$, and $\mathcal{G}_0^T = \mathcal{G}_{A^T}$, where $\mathcal{G}_A$ and $\mathcal{G}_{A^T}$ are the induced graph of A and A^T, respectively.*

We now present the main result on a.s. consensus under independent communication.

Theorem 4 *Suppose A3 (Independent Communication) and A4 (Double Connectivity) hold. Global a.s. consensus is achieved if and only if $\sum_{k=0}^{\infty}(P_k^+ + P_k^-) = \infty$.*

Denote $d_* = \max\left\{\mathrm{diam}(\mathcal{G}_A), \mathrm{diam}(\mathcal{G}_{A^T})\right\}$, where $\mathrm{diam}(\mathcal{G}_A)$ and $\mathrm{diam}(\mathcal{G}_{A^T})$ represent the diameter of the induced graph of A and A^T, respectively. Take $E_* = |\mathcal{E}_0| - \sum_{i=1}^n \mathrm{sgn}(a_{ii})$, where $|\mathcal{E}_0|$ represents the number of elements in $\mathcal{E}_0$, and $\mathrm{sgn}(z)$ is the sign function. Introduce

$a_* = \min\{a_{ij} : a_{ij} > 0, \ i, j = 1, \ldots, n, \ i \neq j\}$ as the lower bound of the nonzero and non-diagonal entries in the meeting probability matrix A. For a.s. ϵ-consensus, the following conclusion holds.

Theorem 5 *Suppose A3 (Independent Communication) and A4 (Double Connectivity) hold. Global a.s. ϵ-consensus is achieved if and only if there exist a constant $p_* > 0$ and an integer $T_* \geq 1$ such that $\sum_{k=m}^{s+T_*-1}(P_k^+ + P_k^-) \geq p_*$ for all $m \geq 0$. In this case, we have*

$$T_{\text{com}}(\epsilon) \leq \frac{4T_*\theta_0/p_*}{\log\left(1 - \left(\frac{a_*}{4n}\right)^{\theta_0}\right)^{-1}} \log \epsilon^{-1} + O(1), \tag{34}$$

where $\theta_0 \doteq (2d_* - 1)(2E_* - 1)$.

We also have the following conclusion indicating that the expected value of the consensus limit ξ equals the initial average x_{ave} if $P_k^+ = P_k^-$.

Theorem 6 *Suppose A3 (Independent Communication) holds and the consensus limit exists. Then $\mathbf{E}(\xi) = x_{\text{ave}}$ if $P_k^+ = P_k^-$ for all k.*

With independent communication, when $P_k^+ = P_k^-$ for all k, it is not hard to see that for any $k = 0, 1, \ldots,$ $\mathbf{E}(W(k))$ is a doubly stochastic matrix since it is both stochastic and symmetric. Thus, Theorem 6 holds trivially. Furthermore, we have another conclusion showing that whenever the consensus limit ξ exists, the average can almost never be preserved.

Theorem 7 *Suppose A3 (Independent Communication) holds and the number of nodes, n, is odd. Assume that $P_k^+, P_k^- \in [0, 1 - \varepsilon]$ for all $k \geq 0$ with $0 < \varepsilon < 1$ a fixed number. Then for any $k_0 \geq 0$ and for almost all initial conditions $x^0 = x(k_0) \in \mathbb{R}^n$, we have*

$$\mathbf{P}\left(\sum_{i=1}^{n} x_i(k) = nx_{\text{ave}}, \quad k \geq k_0\right) = 0 \tag{35}$$

if the consensus limit ξ exists.

Remark 11 *Surprisingly enough Theorem 7 relies on the condition that the number of nodes is odd. This is not conservative in the sense that we can easily find an example with even number of nodes such that $\mathbf{P}\left(\sum_{i=1}^{n} x_i(k) = nx_{\text{ave}}, \ k \geq k_0\right) > 0$ conditioned the consensus limit exists.*

Remark 12 *Note that perfectly dependent communication coincides with independent communication when $P_k^+ = P_k^- = 1$. This is why we need P_k^+ and P_k^- to be at a distance from one in the assumption of Theorem 7.*

Regarding the non-conservativeness of A4 (Double Connectivity) to ensure a consensus under independent communication, we have the following conclusion.

Proposition 1 *Suppose A3 (Independent Communication) holds. Then the condition $\sum_{k=0}^{\infty}(P_k^+ + P_k^-) = \infty$ always ensures an a.s. consensus only if A4 (Double Connectivity) holds.*

Proof. Assume that A4 does not hold. Then either $\mathcal{G}_0$ or $\mathcal{G}_0^T$ is not quasi-strongly connected.

Let us first discuss the case when $\mathcal{G}_0$ is not quasi-strongly connected. There will be two distinct nodes i and j such that $\bar{\mathcal{V}}_1 \cap \bar{\mathcal{V}}_2 = \emptyset$, where $\bar{\mathcal{V}}_1 = \{$nodes from which i is reachable in $\mathcal{G}_0\}$ and $\bar{\mathcal{V}}_2 = \{$nodes from which j is reachable in $\mathcal{G}_0\}$. The definition of the node pair selection process then implies

$$\left\{ m : a_{mk} > 0, k \in \bar{\mathcal{V}}_\tau \right\} \subseteq \bar{\mathcal{V}}_\tau, \quad \tau = 1, 2. \tag{36}$$

Take $P_k^+ = 0$ for all k. Then each node in $\bar{\mathcal{V}}_\tau, \tau = 1, 2$ can only be connected to the nodes in the same subset. Consensus will then fail with probability one even with $\sum_{k=0}^{\infty} P_k^- = \infty$ if we just take $x_m(k_0) = 0$ for $m \in \bar{\mathcal{V}}_1$ and $x_m(k_0) = 1$ for $m \in \bar{\mathcal{V}}_2$.

Similar analysis leads to the same conclusion for the case with $\mathcal{G}_0^T$ not quasi-strongly connected. This completes the proof. □

Remark 13 *It was shown in Remark 5 that A1 (Weak Connectivity) is a lower bound for the underlying connectivity ensuring a consensus, and Theorem 1 indicated that under A1, the condition $\sum_{k=0}^{\infty}(P_k^+ + P_k^-) = \infty$ ensures an a.s. consensus under independent communication. Therefore, combining Remark 5, Theorem 1, Theorem 4, and Proposition 1, we can conclude that in terms of consensus convergence of the randomized gossip algorithm, A1 (Weak Connectivity) is critical for perfectly dependent communication, as is A4 (Double Connectivity) for independent communication.*

Remark 14 *For perfectly dependent communication, Theorem 3 shows the consensus limit equals the initial average with probability one. While for independent communication, we see from Theorem 6 that only the expected value of the consensus limit equals the initial average with an additional condition $P_k^+ = P_k^-$.*

Remark 15 *We see from (12) and (35) that with odd number of nodes, average preserving turns from almost forever (with probability one for all initial conditions) with perfectly dependent*

communication, to almost never (with probability zero for almost all initial conditions) for the independent case. As has been shown widely in the classical random graph theory [6, 8, 7], the probability of a random graph to hold a certain property often jumps from one to zero when a certain threshold is crossed. Now we can conclude that communication dependence provides such a threshold for average preserving of the considered randomized gossip algorithm.

With independent communication, we have $W(k) = I - \frac{e_i(e_i-e_j)^T}{2}$ with a nontrivial probability in (33). The matrix $I - \frac{e_i(e_i-e_j)^T}{2}$ is neither doubly stochastic nor symmetric, and it is not a projection matrix. These properties are critical for the convergence analysis in Theorems 1 and 2 to stand. Hence, we need new methods to analyze the convergence property under assumption A3.

The rest of this section is organized as follows. In Subsection 5.1, we establish several useful lemmas for the convergence analysis. Proofs of Theorems 4, 5 and 7 are given in Subsection 5.2.

5.1 Key Lemmas

In this subsection, we first establish an important property of the Bernoulli trials defined by the node communication process. Then, we investigate the product of the stochastic matrices derived from the gossip algorithm.

5.1.1 Bernoulli Communication Links

Define two (independent) sequences of independent Bernoulli trials

$$\mathfrak{B}_0^+, \mathfrak{B}_1^+, \mathfrak{B}_2^+, \ldots,$$

$$\mathfrak{B}_0^-, \mathfrak{B}_1^-, \mathfrak{B}_2^-, \ldots,$$

such that $\mathbf{P}\big(\mathfrak{B}_k^+ = 1\big) = P_k^+$ and $\mathbf{P}\big(\mathfrak{B}_k^- = 1\big) = P_k^-$. Then let

$$\mathfrak{B}_0, \mathfrak{B}_1, \mathfrak{B}_2, \ldots,$$

denote the independent Bernoulli trials given by $\mathfrak{B}_k = 1$ if and only if $\mathfrak{B}_k^- + \mathfrak{B}_k^+ \geq 1$.

Clearly $\mathfrak{B}_k = 1$ characterizes the event that at least one side of the communication succeeds at time k. In order to establish the success times property of the process of $\{\mathfrak{B}_k\}_0^\infty$, we need the following lemma.

Lemma 4 *Suppose $\{b_k\}, \{c_k\}$ are two sequences satisfying $b_k, c_k \in [0,1]$, $k = 0, 1, \ldots$. Then $\sum_{k=0}^\infty \big(b_k + c_k - b_k c_k\big) = \infty$ if and only if $\sum_{k=0}^\infty \big(b_k + c_k\big) = \infty$.*

Proof. It is straightforward to see that

$$\frac{1}{2}(b_k + c_k) \leq \max\{b_k, c_k\} \leq b_k + c_k - b_k c_k \leq b_k + c_k. \tag{37}$$

Then the desired conclusion follows trivially. □

The following lemma holds on the success times of $\{\mathfrak{B}_k\}_0^\infty$.

Lemma 5 $\mathbf{P}\big(for\ all\ k_0 \geq 0,\ \mathfrak{B}_k = 1\ for\ infinitely\ many\ k \geq k_0\big) = 1$ *if and only if* $\sum_{k=0}^\infty (P_k^+ + P_k^-) = \infty.$

Proof. Note that, we have $\mathbf{P}\big(\mathfrak{B}_k = 1\big) = 1 - (1 - P_k^+)(1 - P_k^-) = P_k^+ + P_k^- - P_k^+ P_k^- \doteq z_k$. Then

$$\mathbf{P}\big(\mathfrak{B}_k = 0, k \geq k_0\big) = \prod_{k=k_0}^\infty (1 - z_k), \tag{38}$$

and the necessity part of the conclusion follows immediately from Lemmas 2 and 4.

On the other hand, since $\sum_{k=0}^\infty (P_k^+ + P_k^-) = \infty$ implies $\prod_{k=s}^\infty (1 - z_k) = 0$ for all $s \geq 0$, we have

$$\begin{aligned}
\mathbf{P}\big(\exists m \geq 0,\ \mathfrak{B}_k = 1 \text{ for finitely many } k \geq m\big) &\leq \sum_{m=0}^\infty \mathbf{P}\big(\mathfrak{B}_k = 1 \text{ for finitely many } k \geq m\big) \\
&\leq \sum_{m=0}^\infty \sum_{s=m}^\infty \mathbf{P}\big(\mathfrak{B}_k = 0 \text{ for } k \geq s\big) \\
&\leq \sum_{m=0}^\infty \sum_{s=m}^\infty \prod_{k=s}^\infty (1 - z_k) \\
&= 0. \tag{39}
\end{aligned}$$

This leads to $\mathbf{P}\big(\text{for all } k_0 \geq 0,\ \mathfrak{B}_k = 1 \text{ for infinitely many } k \geq k_0\big) = 1$. We have now completed the proof. □

5.1.2 Products of Transition Matrices

The considered gossip algorithm is determined by the possible samples of the transition matrix $W(k)$. Denote $\mathbb{M} = \mathbb{M}_1 \bigcup \mathbb{M}_2$ with

$$\mathbb{M}_1 = \left\{ I - \frac{e_i(e_i - e_j)^T}{2} : \quad a_{ij} + a_{ji} > 0,\ i, j = 1, \ldots, n,\ i \neq j \right\}$$

and

$$\mathbb{M}_2 = \left\{ I - \frac{(e_i - e_j)(e_i - e_j)^T}{2} : \quad a_{ij} + a_{ji} > 0,\ i, j = 1, \ldots, n,\ i > j \right\}.$$

Then $\mathbb{M}$ is the set including all samples of $W(k)$ except for the identity matrix I. The following lemma holds on the product of matrices from $\mathbb{M}$.

20

Lemma 6 *Let $M_k \in \mathbb{M}$, $k = 1, \dots, N$ be $N \geq 1$ matrices in $\mathbb{M}$. Then we have $\left(\bigcup_{i=1}^{N} \mathcal{G}_{M_i}\right) \subseteq \mathcal{G}_{M_N \cdots M_1}$. Moreover, all nonzero entries of the product $M_N \cdots M_1$ have lower bound 2^{-N}.*

Proof. We just prove the case for $N = 2$. Then the conclusion follows immediately by induction.

Note that the nonzero entries of any matrix in $\mathbb{M}$ is no smaller than $1/2$. Moreover, all matrices in $\mathbb{M}$ have positive diagonal entries. Therefore, denoting $\bar{m}_{ij}$, $\hat{m}_{ij}$ and m_{ij}^* as the ij-entries of M_1, M_2 and $M_1 M_2$, respectively, we have

$$m_{i_1 i_2}^* = \sum_{j=1}^{n} \bar{m}_{i_1 j} \hat{m}_{j i_2} \geq \bar{m}_{i_1 i_2}/2 + \bar{m}_{i_1 i_1}/2. \tag{40}$$

This implies $m_{i_1 i_2}^* > 0$ as long as at least one of $\bar{m}_{i_1 i_2}$ and $\hat{m}_{i_1 i_2}$ is non-zero. Furthermore, if $m_{i_1 i_2}^* > 0$, it is straightforward to see that $m_{i_1 i_2}^* \geq 1/4$. This completes the proof. $\qquad \square$

Recall that $d_* = \max\left\{\operatorname{diam}(\mathcal{G}_A), \operatorname{diam}(\mathcal{G}_{A^{\mathrm{T}}})\right\}$, where $\operatorname{diam}(\mathcal{G}_A)$ and $\operatorname{diam}(\mathcal{G}_{A^{\mathrm{T}}})$ represent the diameter of the induced graph of A and A^T, respectively. We have the following lemma.

Lemma 7 *Suppose A4 (Double Connectivity) holds. Let $M_1, \dots, M_{2d_*-1}$ be $2d_* - 1$ products of some finite matrices in $\mathbb{M}$. If $\mathcal{G}_A \subseteq \mathcal{G}_{M_k}$ or $\mathcal{G}_{A^T} \subseteq \mathcal{G}_{M_k}$ holds for any $k = 1, \dots, 2d_* - 1$, then $M_{2d_*-1} \cdots M_1$ is a scrambling matrix.*

Proof. Since for any k, we have $\mathcal{G}_A \subseteq \mathcal{G}_{M_k}$ or $\mathcal{G}_{A^T} \subseteq \mathcal{G}_{M_k}$, there must be one of $\mathcal{G}_A \subseteq \mathcal{G}_{M_k}$ or $\mathcal{G}_{A^T} \subseteq \mathcal{G}_{M_k}$ happens at least d_* times for $k = 1, \dots, 2d_* - 1$. Without loss of generality, we just focus on the case that $\mathcal{G}_A \subseteq \mathcal{G}_{M_k}$ happens at least d_* times. Thus, there exist $1 \leq k_1 < k_2 < \cdots < k_{d_*} \leq 2d_* - 1$ such that $\mathcal{G}_A \subseteq \mathcal{G}_{M_{k_s}}$, $s = 1, \dots, d_*$. We now separate the product $M_{2d_*-1} \cdots M_1$ into the product of d_* matrices:

$$\bar{M}_1 = M_{k_1} \cdots M_1; \quad \bar{M}_j = M_{k_j} \cdots M_{k_{j-1}+1}, j = 2, \dots, d_* - 1; \quad \bar{M}_{d_*} = M_{2d_*-1} \cdots M_{k_{d_*-1}+1}.$$

Then $M_{2d_*-1} \cdots M_1 = \bar{M}_{d_*} \cdots \bar{M}_1$. Since $\mathcal{G}_A \subseteq \mathcal{G}_{M_{k_s}}$ for each k_s, we have $\mathcal{G}_A \subseteq \mathcal{G}_{M_{k_s}} \subseteq \mathcal{G}_{\bar{M}_s}$, $s = 1, \dots, d_*$ directly from Lemma 6.

Suppose i_0 is a center node of $\mathcal{G}_A$. Then for any other node j_0, there must be a path $i_0 \to j_0$ in $\mathcal{G}_A$ with length no larger than d_*. We assume $d(i_0, j_0) = d_0 \leq d_*$. We redenote j_0 as i_{d_0}, and let $i_0 e_0 i_1 \dots e_{d_0} i_{d_0}$ be a shortest path from i_0 to i_{d_0} in the graph $\mathcal{G}_A$.

We denote the ij-entry of $\bar{M}_s$ as $\bar{M}_{ij}^{\langle s \rangle}$ for $s = 1, \dots, d_*$. Also denote the ij-entry of $\bar{M}_2 \bar{M}_1$ as $\bar{M}_{ij}^{\langle 2 \rangle \langle 1 \rangle}$. The fact that $(i_0, i_1) \in \mathcal{E}_A \subseteq \mathcal{E}_{\bar{M}_1}$ immediately implies $\bar{M}_{i_1 i_0}^{\langle 1 \rangle} > 0$ according to the definition of induced graph. Similarly we have $\bar{M}_{i_2 i_1}^{\langle 2 \rangle} > 0$ because $(i_1, i_2) \in \mathcal{E}_A \subseteq \mathcal{E}_{\bar{M}_2}$. Thus, we

obtain

$$\bar{M}_{i_2 i_0}^{\langle 2 \rangle \langle 1 \rangle} = \sum_{\tau=1}^{n} \bar{M}_{i_2 \tau}^{\langle 2 \rangle} \bar{M}_{\tau i_0}^{\langle 1 \rangle} \geq \bar{M}_{i_2 i_1}^{\langle 2 \rangle} \bar{M}_{i_1 i_0}^{\langle 1 \rangle} > 0. \tag{41}$$

Similar analysis can be proceeded until we eventually obtain

$$\bar{M}_{i_{d_0} i_0}^{\langle d_0 \rangle \ldots \langle 1 \rangle} > 0 \tag{42}$$

where $\bar{M}_{i_\tau i_0}^{\langle d_0 \rangle \ldots \langle 1 \rangle}$ denotes the $i_\tau i_0$-entry of $\bar{M}_{d_0} \cdots \bar{M}_1$.

Denote the τi_0-entry of $M_{2d_*-1} \ldots M_1$ as $M_{\tau i_0}^*$. Noting that (42) holds for arbitrary i_{d_0}, we see from Lemma 6 that

$$M_{\tau i_0}^* = \bar{M}_{\tau i_0}^{\langle d_* \rangle \ldots \langle 1 \rangle} > 0, \ \tau = 1, \ldots, n. \tag{43}$$

Therefore, according to the definition of $\lambda(\cdot)$ in (1), obtain

$$\lambda \big(M_{2d_*-1} \cdots M_1 \big) \leq 1 - \min_{j=1,\ldots,n} M_{j i_0}^* < 1.$$

We have now proved the conclusion. $\qquad \square$

We further denote $\mathbb{M}_2^* = \left\{ I - \frac{(e_i - e_j)(e_i - e_j)^T}{2} : \ i, j = 1, \ldots, n, \ i > j \right\}$. The following lemma is on the impossibility of the finite-time convergence for the product of matrices in $\mathbb{M}_2^*$.

Lemma 8 *Suppose n is an odd number. Take matrices $M_\tau \in \mathbb{M}_2^*$, $\tau = 1, \ldots, k, k \geq 1$ arbitrarily. Then we have $\delta(M_k \cdots M_1) > 0$, where $\delta(\cdot)$ is defined in (1).*

Proof. We prove the lemma by a contradiction argument. Suppose there exist an integer $k_* \geq 1$ and k_* matrices $M_\tau \in \mathbb{M}_2^*, \tau = 1, \ldots, k_*$ satisfying $\delta(M_{k_*} \cdots M_1) = 0$. Therefore, there exists $\beta \in \mathbb{R}^n$ as an $n \times 1$ vector such that

$$M_{k_*} \cdots M_1 = \mathbf{1} \beta^T, \tag{44}$$

which implies

$$M_{k_*} \cdots M_1 y = \mathbf{1} \beta^T y = z_0 \mathbf{1} \tag{45}$$

for all $y = (y_1 \ldots y_n)^T \in \mathbb{R}^n$, where $z_0 = \beta^T y$ is a scalar. Since $M_\tau \in \mathbb{M}_2^*, \tau = 1, \ldots, k_*$, the average of $y_1, \ldots, y_n$ is always preserved, and therefore, $z_0 = \sum_{i=1}^{n} y_i / n$.

Let $n = 2n_0 + 1$ since n is an odd number. Take $y_i = 0$ for $i = 1, \ldots, n_0$ and $y_i = 2^{k_*+1}$ for $i = n_0 + 1, \ldots, 2n_0 + 1$. Then $z_0 = 2^{k_*+1}(n_0 + 1)/(2n_0 + 1)$. On the other hand, it is straightforward to verify that each element in $M_{k_*} \cdots M_1 y$ can only be an even number, say, S_0.

Since S_0 is even, we have $S_0 = 2^a S_*$ for some $0 \leq a \leq k_*$ an integer and S_* an odd number. This leads to

$$2^{k_*+1} \frac{n_0 + 1}{2n_0 + 1} = 2^a S_*, \tag{46}$$

which implies

$$2^{k_*+1-a}(n_0 + 1) = (2n_0 + 1)S_*. \tag{47}$$

Clearly it is impossible for (47) to hold true since its left-hand side is an even number, while the right-hand side odd. Therefore such $M_\tau \in \mathbb{M}_2^*, \tau = 1, \ldots, k_*$ does not exist and the desired conclusion follows. $\qquad\square$

5.2 Proofs

We are now in a place to prove Theorems 4, 5 and 7.

5.2.1 Proof of Theorem 4

(Necessity.) We have

$$
\begin{aligned}
\mathbf{P}\big(x_i(k+1) = x_i(k)\big) &\geq 1 - \sum_{j=1,\ j\neq i} \Big[\mathbf{P}\big(\text{pair } (j,i) \text{ is selected}\big) \cdot \mathbf{P}\big(i \text{ receives } x_j(k)\big) \\
&\qquad + \mathbf{P}\big(\text{pair } (i,j) \text{ is selected}\big) \cdot \mathbf{P}\big(i \text{ receives } x_j(k)\big)\Big] \\
&= 1 - P_k^+ \sum_{j=1,\ j\neq i}^{n} \frac{a_{ij}}{n} - P_k^- \sum_{j=1,\ j\neq i}^{n} \frac{a_{ji}}{n} \\
&\geq 1 - h_i \max\big\{P_k^+, P_k^-\big\}
\end{aligned}
\tag{48}
$$

where h_i is introduced in (21). Note that, from (37), we know $\sum_{k=0}^{\infty}(P_k^+ + P_k^-) = \infty$ if and only if $\sum_{k=0}^{\infty} \max\big\{P_k^+, P_k^-\big\} = \infty$. Therefore, with (48), the necessity part of Theorem 4 follows immediately from the same argument as the proof of the the necessity part in Theorem 1.

(Sufficiency.) Recall that the considered gossip algorithm is determined by the random matrix, $W(k)$ in (33). Denote the induced (random) graph of $W(k)$ as $\mathcal{G}_{W(k)} = \big(\mathcal{V}, \mathcal{E}_{W(k)}\big)$. Global a.s. consensus is equivalent to $\mathbf{P}\big(\lim_{k\to\infty} \delta(W_k \cdots W_2 W_1) = 0\big) = 1$.

With independent communication, we have

$$\mathbf{P}\Big((i,j) \in \mathcal{E}_{W(k)} \Big| \text{at least one of } \mathfrak{B}_k^+ \text{ and } \mathfrak{B}_k^- \text{ succeeds at time } k\Big)$$

$$= \frac{(a_{ji}P_k^+ + a_{ij}P_k^-)/n}{1 - (1 - P_k^+)(1 - P_k^-)}$$

$$= \frac{a_{ji}P_k^+ + a_{ij}P_k^-}{n(P_k^+ + P_k^- - P_k^+P_k^-)}$$

$$\geq \frac{a_{ji}P_k^+ + a_{ij}P_k^-}{n(P_k^+ + P_k^-)} \tag{49}$$

for all $i, j = 1, \ldots, n, \ i \neq j$. Here without loss of generality, we assume $P_k^+ + P_k^- > 0$.

Recall that $a_* = \min\{a_{ij} : a_{ij} > 0, \ i, j = 1, \ldots, n, \ i \neq j\}$ is the lower bound of the nonzero and non-diagonal entries in the meeting probability matrix A. Based on (49), there are two cases.

(i) When $P_k^+ \geq P_k^-$, for all $i, j = 1, \ldots, n, \ i \neq j$ with $a_{ij} > 0$, we have

$$\mathbf{P}\Big((j,i) \in \mathcal{E}_{W(k)} \Big| \text{at least one of } \mathfrak{B}_k^+ \text{ and } \mathfrak{B}_k^- \text{ succeeds at time } k\Big) \geq \frac{a_*}{2n}. \tag{50}$$

(i) When $P_k^+ < P_k^-$, for all $i, j = 1, \ldots, n, \ i \neq j$ with $a_{ij} > 0$, we have

$$\mathbf{P}\Big((i,j) \in \mathcal{E}_{W(k)} \Big| \text{at least one of } \mathfrak{B}_k^+ \text{ and } \mathfrak{B}_k^- \text{ succeeds at time } k\Big) \geq \frac{a_*}{2n}. \tag{51}$$

Now we introduce the Bernoulli (success) sequence of $\mathfrak{B}_0, \mathfrak{B}_1, \mathfrak{B}_2, \ldots$ as

$$0 \leq \zeta_1 < \zeta_2 < \cdots : \quad \mathfrak{B}_{\zeta_m} = 1$$

where ζ_m is the time of the m'th success. Since $\sum_{k=0}^{\infty}(P_k^+ + P_k^-) = \infty$, Lemma 5 guarantees that $\zeta_m < \infty$ a.s. for all $m = 1, 2, \ldots$. For simplicity, we assume $k_0 = 0$ in the rest of the proof.

From (50) and (51), for $\zeta_1 < \zeta_2 < \ldots$, we can take a sequence of arcs $(i_1, j_1), (i_2, j_2), \ldots$ such that

$$\mathbf{P}\Big((i_m, j_m) \in \mathcal{E}_{W(\zeta_m)}\Big) \geq \frac{a_*}{2n}, \tag{52}$$

where $i_m \neq j_m$ for all $m = 1, 2, \ldots$ and either $a_{i_m j_m} > 0$ or $a_{j_m i_m} > 0$ holds. Moreover, since the node pair selection process and the node communication process are independent for different instances, events $\{(i_m, j_m) \in \mathcal{E}_{W(\zeta_m)}, m = 1, 2, \ldots\}$ are independent.

Recall that E_* is the number of non-self-looped arcs in the underlying graph. From the double connectivity assumption A4, it is not hard to see that we can select $(i_1, j_1), (i_2, j_2), \ldots, (i_{2E_*-1}, j_{2E_*-1})$ properly such that

$$\mathcal{G}_A \subseteq \bigcup_{m=1}^{2E_*-1} \{(i_m, j_m)\} \quad \text{or} \quad \mathcal{G}_{A^T} \subseteq \bigcup_{m=1}^{2E_*-1} \{(i_m, j_m)\}$$

holds. Thus, based on Lemma 6, denoting $Q_1 = W_{\zeta_{2E_*-1}} \cdots W_{\zeta_1}$, we have $\mathcal{G}_A \subseteq \mathcal{G}_{Q_1}$ or $\mathcal{G}_{A^T} \subseteq \mathcal{G}_{Q_1}$.

Similarly, denoting $Q_\tau = W_{\zeta_{(2E_*-1)\tau}} \cdots W_{\zeta_{(2E_*-1)(\tau-1)+1}}$ for $\tau = 1, 2, \ldots$, we have $\mathcal{G}_A \subseteq \mathcal{G}_{Q_\tau}$ or $\mathcal{G}_{A^T} \subseteq \mathcal{G}_{Q_\tau}$ for all τ. According to Lemma 7, we have

$$\mathbf{P}\Big(\lambda(Q_{2d_*-1} \cdots Q_1) < 1\Big) \geq \Big(\frac{a_*}{2n}\Big)^{(2d_*-1)(2E_*-1)}. \tag{53}$$

Moreover, since $Q_{2d_*-1} \ldots Q_1$ is a product of $\theta_0 = (2d_* - 1)(2E_* - 1)$ matrices in $\mathbb{M}$, Lemma 6 further ensures

$$\mathbf{P}\Big(\lambda(Q_{2d_*-1} \cdots Q_1) < 1 - 2^{-\theta_0}\Big) \geq \Big(\frac{a_*}{2n}\Big)^{\theta_0}. \tag{54}$$

Continuing the analysis, we know that for all $F_s = Q_{(2d_*-1)s} \cdots Q_{(2d_*-1)(s-1)+1}$, $s = 1, 2, \ldots$, we have

$$\mathbf{P}\Big(\lambda(F_s) < 1 - 2^{-\theta_0}\Big) \geq \Big(\frac{a_*}{2n}\Big)^{\theta_0}, \quad s = 1, 2, \ldots, \tag{55}$$

which implies

$$\mathbf{E}\Big(\lambda(F_s)\Big) \leq 1 - \Big(\frac{a_*}{4n}\Big)^{\theta_0}, \quad s = 1, 2, \ldots. \tag{56}$$

With Fatou's lemma and Lemma 1, we finally have

$$\mathbf{E}\Big(\lim_{m\to\infty} \delta\big(F_m \cdots F_1\big)\Big) \leq \lim_{m\to\infty} \mathbf{E}\Big(\delta\big(F_m \cdots F_1\big)\Big) \leq \lim_{m\to\infty} \mathbf{E}\Big(\prod_{s=1}^{m} \lambda(F_s)\Big) = 0, \tag{57}$$

which implies

$$\mathbf{P}\Big(\lim_{m\to\infty} \delta\big(F_m \cdots F_1\big) = 0\Big) = 1. \tag{58}$$

Note that, (58) leads to $\mathbf{P}\big(\lim_{k\to\infty} \delta(W_k \cdots W_1) = 0\big) = 1$ since the definition of F_s, $s = 1, 2, \ldots$ guarantees $W_k = I$ for all $k \notin \{\zeta_1, \zeta_2, \ldots\}$. This completes the proof.

5.2.2 Proof of Theorem 5

Similar to the proof of Theorem 4, the necessity part of Theorem 5 follows from the same argument as the one used in the proof of Theorem 2. Here we just focus on the sufficiency statement of Theorem 5. For simplicity we assume $k_0 = 0$.

Denote the ij-entry of $W(k-1)\cdots W(0)$ as $\Psi_{ij}(k)$. Then for all i, j and α, we have

$$
\begin{aligned}
\big|x_i(k) - x_j(k)\big| &= \Big|\sum_{m=1}^{n} \Psi_{im}(k)x_m(0) - \sum_{m=1}^{n} \Psi_{jm}(k)x_m(0)\Big| \\
&= \Big|\sum_{m=1}^{n} \Psi_{im}(k)\big(x_m(0) - x_\alpha(0)\big) - \sum_{m=1}^{n} \Psi_{jm}(k)\big(x_m(0) - x_\alpha(0)\big)\Big| \\
&\leq \sum_{m=1}^{n} \big|\Psi_{im}(k) - \Psi_{jm}(k)\big| \cdot \max_{m}\big|x_m(0) - x_\alpha(0)\big| \\
&\leq n\delta\big(W(k-1)\cdots W(0)\big)\mathcal{H}(0),
\end{aligned}
\tag{59}
$$

which implies

$$
\mathcal{H}(k) \leq n\delta\big(W(k-1)\cdots W(0)\big)\mathcal{H}(0).
\tag{60}
$$

Then we introduce $\rho_k = \sum_{m=0}^{k-1}\mathfrak{B}_m$. From Markov's inequality, we have

$$
\begin{aligned}
\mathbf{P}\Big(\frac{\mathcal{H}(k)}{\mathcal{H}(0)} \geq \epsilon\Big) &\leq \mathbf{P}\Big(\delta\big(W(k-1)\cdots W(0)\big) \geq \frac{\epsilon}{n}\Big) \\
&\leq \frac{n}{\epsilon}\mathbf{E}\Big(\delta\big(W(k-1)\cdots W(0)\big)\Big) \\
&= \frac{n}{\epsilon}\mathbf{E}\Big(\delta\big(W_{\zeta_{\rho_k}}\cdots W_{\zeta_1}\big)\Big) \\
&\leq \frac{n}{\epsilon}\mathbf{E}\Big(\lambda\big(F_{\lfloor\frac{\rho_k}{\theta_0}\rfloor}\big)\cdots\lambda\big(F_1\big)\Big),
\end{aligned}
\tag{61}
$$

where θ_0 and $F_m, m = 1, 2, \ldots$ are defined in the proof of Theorem 4. Thus, (56) and (61) lead to

$$
\begin{aligned}
\mathbf{E}\Big(\lambda\big(F_{\lfloor\frac{\rho_k}{\theta_0}\rfloor}\big)\cdots\lambda\big(F_1\big)\Big) &= \mathbf{E}\Big(\mathbf{E}\Big(\lambda\big(F_{\lfloor\frac{\rho_k}{\theta_0}\rfloor}\big)\cdots\lambda\big(F_1\big)\big|\rho_k\Big)\Big) \\
&= \Big(1 - \big(\frac{a_*}{4n}\big)^{\theta_0}\Big)^{\mathbf{E}\left(\lfloor\frac{\rho_k}{\theta_0}\rfloor\right)} \\
&\leq \Big(1 - \big(\frac{a_*}{4n}\big)^{\theta_0}\Big)^{\frac{\mathbf{E}(\rho_k)}{\theta_0}-1} \\
&= \Big(1 - \big(\frac{a_*}{4n}\big)^{\theta_0}\Big)^{\frac{\sum_{m=0}^{k-1}\left(P_k^+ + P_k^- - P_k^+ P_k^-\right)}{\theta_0}-1} \\
&\leq \Big(1 - \big(\frac{a_*}{4n}\big)^{\theta_0}\Big)^{\frac{\sum_{m=0}^{k-1}\left(P_k^+ + P_k^-\right)}{2\theta_0}-1}
\end{aligned}
\tag{62}
$$

since both the node pair selection process and the node communication process are independent for different instances.

Since there exist a constant $p_* > 0$ and an integer $T_* \geq 1$ such that $\sum_{k=m}^{s+T_*-1}(P_k^+ + P_k^-) \geq p_*$ for all $m \geq 0$, (61) and (62) imply

$$\mathbf{P}\Big(\frac{\mathcal{H}(k)}{\mathcal{H}(0)} \geq \epsilon\Big) \leq \frac{n}{\epsilon}\Big(1 - \Big(\frac{a_*}{4n}\Big)^{\theta_0}\Big)^{\frac{\lfloor\frac{k}{T_*}\rfloor p_*}{2\theta_0}-1} \tag{63}$$

and thus

$$T_{\mathrm{com}}(\epsilon) \leq \frac{4T_*\theta_0/p_*}{\log\big(1 - \big(\frac{a_*}{4n}\big)^{\theta_0}\big)^{-1}}\log\epsilon^{-1} + O(1). \tag{64}$$

The desired conclusion follows.

5.2.3 Proof of Theorem 7

With independent communication, it follows from Theorem 4 that $\sum_{k=0}^{\infty}\big(P_k^+ + P_k^-\big) = \infty$ if the consensus limit ξ exists. It is not hard to see that

$$\begin{aligned}
\mathbf{P}\big(W(k) \in \mathbb{M}_1\big) &= \Big(1 - \frac{\sum_{i=1}^{n} a_{ii}}{n}\Big)\Big(P_k^+\big(1 - P_k^-\big) + P_k^-\big(1 - P_k^+\big)\Big) \\
&= \Big(1 - \frac{\sum_{i=1}^{n} a_{ii}}{n}\Big)\big(P_k^+ + P_k^- - 2P_k^+ P_k^-\big).
\end{aligned} \tag{65}$$

Since $P_k^+, P_k^- \in [0, 1-\varepsilon]$ for all $k \geq 0$ with $0 < \varepsilon < 1$, we have

$$\begin{aligned}
P_k^+ + P_k^- - 2P_k^+ P_k^- &= P_k^+ + P_k^- - P_k^+ P_k^- - P_k^+ P_k^- \\
&\geq \max\{P_k^+, P_k^-\} - (1-\varepsilon)\max\{P_k^+, P_k^-\} \\
&= \varepsilon\max\{P_k^+, P_k^-\} \\
&\geq \frac{\varepsilon}{2}\big(P_k^+ + P_k^-\big).
\end{aligned} \tag{66}$$

As a result, we have $\sum_{k=0}^{\infty}\big(P_k^+ + P_k^- - 2P_k^+ P_k^-\big) = \infty$. By a similar argument as we obtain Lemma 5, for any $k_0 \geq 0$, we have

$$\mathbf{P}\Big(W(k) \in \mathbb{M}_1 \text{ for infinitely many } k \text{ with } k \geq k_0\Big) = 1 \tag{67}$$

conditioned that the consensus limit ξ exists.

Then we show that for almost all initial conditions, it is impossible to generate finite-time convergence along every sample path $\{W^\omega(k)\}_0^\infty$ of the random matrix process $\{W(k)\}_0^\infty$ which satisfies $W^\omega(k) \in \mathbb{M}_2$ for all k. For any $k = 1, 2, \ldots$, we define

$$\mathcal{I}_k = \big\{x^0 \in \mathbb{R}^n : \exists z \in \mathbb{R}, \Gamma_i \in \mathbb{M}_2, i = 1, \ldots, k \text{ s.t. } \Gamma_k \cdots \Gamma_1 x^0 = z\mathbf{1}\big\}.$$

27

Suppose $\Gamma_1, \ldots, \Gamma_k \in \mathbb{M}_2$. We denote

$$\Gamma_k \cdots \Gamma_1 = \left(\Phi_1 \ldots \Phi_n\right)^T \tag{68}$$

where Φ_m is the m'th column of $\Gamma_k \cdots \Gamma_1$. With Lemma 8, we know that $\delta(\Gamma_k \cdots \Gamma_1) > 0$, which implies $\bigcap_{i=1}^n \left(\text{span}\{\Phi_i\}\right)^\perp$ is a linear space with dimension no larger than $n - 2$ noticing that $\Gamma_k \cdots \Gamma_1$ is a stochastic matrix. Therefore, since $\Gamma_k \cdots \Gamma_1 x^0 = z\mathbf{1}$ leads to $\Gamma_k \ldots \Gamma_1\left(x^0 - z\mathbf{1}\right) = 0$, we have

$$\mathcal{I}_k = \bigcup_{\Gamma_1, \ldots, \Gamma_k \in \mathbb{M}_2} \mathbb{R}^1 \times \mathcal{Y}_{\Gamma_k \cdots \Gamma_1},$$

where $\mathcal{Y}_{\Gamma_k \cdots \Gamma_1} \doteq \left\{y \in \mathbb{R}^n : \Gamma_k \cdots \Gamma_1 y = 0\right\} = \bigcap_{i=1}^n \left(\text{span}\{\Phi_i\}\right)^\perp$. Noting the fact that $\mathbb{M}_2$ is a finite set, we further conclude that $\mathbf{Me}(\mathcal{I}_k) = 0$, where $\mathbf{Me}(S)$ for $S \subseteq \mathbb{R}^n$ denotes the standard Lebesgue measure in $\mathbb{R}^n$. This immediately implies

$$\mathbf{Me}\left(\bigcup_{k=1}^\infty \mathcal{I}_k\right) = 0. \tag{69}$$

Now we observe that if $W^\omega(k) = I - \frac{e_u(e_u - e_v)^T}{2}$ for some $u, v \in \mathcal{V}, u \neq v$ and $k \geq k_0$, $\sum_{i=1}^n x_i(k) = nx_{\text{ave}}$ implies $x_u^\omega(k) = x_v^\omega(k)$. Therefore, in this case $W^\omega(k)$ can be replaced by I without changing the value of $x^\omega(k + 1)$.

Since the node pair selection process and the node communication process are independent with the nodes' states, we conclude from (67) and (69) that

$$\mathbf{P}\left(\sum_{i=1}^n x_i(k) = nx_{\text{ave}}, \quad k \geq k_0 \text{ and consensus limit } \xi \text{ exists}\Big|\xi \text{ exists}\right)$$

$$= \mathbf{P}\left(\sum_{i=1}^n x_i(k) = nx_{\text{ave}}, \quad k \geq k_0 \text{ and consensus achieved in infinite time}\Big|\xi \text{ exists}\right)$$

$$+ \mathbf{P}\left(\sum_{i=1}^n x_i(k) = nx_{\text{ave}}, \quad k \geq k_0 \text{ and consensus achieved in finite time}\Big|\xi \text{ exists}\right)$$

$$\leq \mathbf{P}\left(x_i(k) = x_j(k) \text{ whenever } W(k) = I - \frac{e_i(e_i - e_j)^T}{2} \in \mathbb{M}_1, k \geq k_0 \text{ and}\right.$$

$$\left.\text{consensus achieved in infinite time}\Big|\xi \text{ exists}\right)$$

$$+ \sum_{m=0}^\infty \mathbf{P}\left(W(\tau) \in \mathbb{M}_2, \ \tau = k_0, \ldots, k_0 + m \text{ s.t. } W(k_0 + m) \cdots W(k_0)x^0 = z\mathbf{1}\Big|\xi \text{ exists}\right)$$

$$= 0 + \sum_{m=0}^\infty 0$$

$$= 0 \tag{70}$$

for all $x^0 \in \mathbb{R}^n$ except for $\bigcup_{k=1}^{\infty} \mathcal{I}_k$, which is a set with measure zero.

The desired conclusion follows immediately.

6 Conclusions

This paper presented new results on the role of unreliable node communication in the convergence of randomized gossip algorithms. The model for the random node pair selection process is defined by a stochastic matrix which characterizes the interactions among the nodes in the network. A pair of nodes meets at a random instance, and two Bernoulli communication links are then established between the nodes. Communication on each link succeeds with a time-dependent probability. We presented a series of necessary and sufficient conditions on the success probability sequence to ensure a.s. consensus or ϵ-consensus under perfectly dependent and independent communication processes, respectively. The results showed that the communication symmetry is critical for the convergence.

The results are summarized in the following table. We notice the following characteristics:

	Perfectly Dependent Communication	Independent Communication
Critical Underlying Connectivity	*Weak Connectivity*	*Double Connectivity*
Consensus Limit = Initial Average	*Almost Surely*	*In Expectation*
Average Preserved	*Almost Forever (for all **n**)*	*Almost Never (for odd **n**)*

Figure 1: Summary of the properties of the random gossip algorithms considered in the paper. Perfectly dependent and independent communication gives drastically different behavior.

- In terms of consensus convergence of the randomized gossip algorithm, A1 (Weak Connectivity) is critical for perfectly dependent communication, as is A4 (Double Connectivity) for independent communication.

- For perfectly dependent communication, the consensus limit equals the initial average

with probability one. While for independent communication, only the expected value of the consensus limit equals the initial average for the special case $P_k^+ = P_k^-$.

- Average is preserved almost forever (with probability one for all initial conditions) with perfectly dependent communication, and it is preserved almost never (with probability zero for almost all initial conditions) with independent communication if the number of nodes is odd.

The results illustrate that convergence behavior of distributed algorithms may heavily depend on the probabilistic dependence properties in the information flow.

References

[1] D. P. Bertsekas and J. N. Tsitsiklis. *Introduction to Probability.* Athena Scientific, Massachusetts, 2002.

[2] G. Latouche, V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling.* 1st edition, ASA SIAM, 1999.

[3] C. Godsil and G. Royle. *Algebraic Graph Theory.* New York: Springer-Verlag, 2001.

[4] C. Berge and A. Ghouila-Houri. *Programming, Games, and Transportation Networks.* John Wiley and Sons, New York, 1965.

[5] A. Benveniste, M. Métivier and P. Priouret. *Adaptive Algorithms and Stochastic Approximations.* Springer-Verlag: Berlin, 1990.

[6] B. Bollobás. *Random Graphs.* Cambridge University Press, second edition, 2001.

[7] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming*, 547-566, 1998

[8] P. Erdös and A. Rényi, "On the evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 17-61, 1960.

[9] R. Karp, C. Schindelhauer, S. Shenker, and B. Vcking, "Randomized rumor spreading," in *Proc. Symp. Foundations of Computer Science*, pp. 564-574, 2000.

[10] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. Conf. Foundations of Computer Science*, pp. 482-491, 2003.

[11] S. Boyd, P. Diaconis and L. Xiao, "Fastest mixing markov chain on a graph," *SIAM Review*, Vol. 46, No. 4, pp. 667-689, 2004.

[12] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2508-2530, 2006.

[13] D. Mosk-Aoyama and D. Shah, "Fast distributed algorithms for computing separable functions," *IEEE Transactions on Information Theory*, vol.55, no.7, pp. 2997-3007, 2008

[14] D. Shah, "Gossip Algorithms," *Foundations and Trends in Networking*, Vol. 3, No. 1, pp. 1-125, 2008.

[15] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," *Proc. Amer. Math. Soc.*, vol. 15, pp. 733-736, 1963.

[16] J. Hajnal, "Weak ergodicity in non-homogeneous markov chains," *Proc. Cambridge Philos. Soc.*, no. 54, pp. 233-246, 1958.

[17] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118-121, 1974.

[18] S. Muthukrishnan, B. Ghosh, and M. Schultz, "First and second order diffusive methods for rapid, coarse, distributed load balancing," *Theory of Computing Systems*, vol. 31, pp. 331-354, 1998.

[19] R. Diekmann, A. Frommer, and B. Monien, "Efficient schemes for nearest neighbor load balancing," *Parallel Computing*, vol. 25, pp. 789-812, 1999.

[20] S. Martinez, J. Cortés, and F. Bullo, "Motion coordination with distributed information," *IEEE Control Systems Magazine*, vol. 27, no. 4, pp. 75-88, 2007.

[21] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, pp. 803-812, 1986.

[22] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom.Control*, vol. 48, no. 6, pp. 988-1001, 2003.

[23] R. Olfati-Saber and R. Murray, "Consensus problems in the networks of agents with switching topology and time dealys," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520-1533, 2004.

[24] J. Fax and R. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1465-1476, 2004.

[25] M. Cao, A. S. Morse and B. D. O. Anderson, "Reaching a consensus in a dynamically changing environment: a graphical approach," *SIAM J. Control Optim.*, vol. 47, no. 2, 575-600, 2008.

[26] M. Cao, A. S. Morse and B. D. O. Anderson, "Reaching a consensus in a dynamically changing environment: convergence rates, measurement delays, and asynchronous events," *SIAM J. Control Optim.*, vol. 47, no. 2, 601-623, 2008.

[27] M. Cao, A. S. Morse and B. D. O. Anderson, "Agreeing asynchronously," *IEEE Trans. Autom. Control*, vol. 53, no. 8, 1826-1838, 2008.

[28] W. Ren and R. Beard, "Consensus seeking in multi-agent systems under dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, no. 5, pp. 655-661, 2005.

[29] L. Moreau, "Stability of multi-agent systems with time-dependent communication links," *IEEE Trans. Autom. Control*, vol. 50, pp. 169-182, 2005.

[30] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Trans. on Autom. Control*, vol. 50, no. 11, pp. 1867-1872, 2005.

[31] C. W. Wu, "Synchronization and convergence of linear dynamics in random directed networks," *IEEE Trans. Autom. Control*, vol. 51, no. 7, pp. 1207-1210, 2006.

[32] A. Tahbaz-Salehi and A. Jadbabaie, "A necessary and sufficient condition for consensus over random networks," *IEEE Trans. on Autom. Control*, VOL. 53, NO. 3, pp. 791-795, 2008.

[33] F. Fagnani and S. Zampieri, "Asymmetric randomized gossip algorithms for consensus," *IFAC World Congress*, Seoul, pp. 9051-9056, 2008.

[34] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE J. on Selected Areas of Communications*, vol. 26, no.4, pp. 634-649, 2008.

[35] F. Fagnani and S. Zampieri, "Average consensus with packet drop communication," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 102-133, 2009.

[36] S. Patterson, B. Bamieh and A. El Abbadi, "Convergence rates of distributed average consensus with stochastic link failures," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 880-892, 2010.

[37] I. Matei, N. Martins and J. S. Baras, "Almost sure convergence to consensus in markovian random graphs," in *Proc. IEEE Conf. Decision and Control*, pp. 3535-3540, 2008.

[38] C. C. Moallemi and B. Van Roy, "Consensus propagation," *IEEE Trans. Information Theory*, vol. 52, no. 11, pp. 4753-4766, 2006.

[39] T. C. Aysal and K. E. Barner, "Convergence of consensus models with stochastic disturbances," *IEEE Trans. Information Theory*, vol. 56, no. 8, pp. 4101-4113, 2010.

[40] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: quantized data and random link failures," *IEEE Trans. signal Processing*, Vol. 58:3, pp. 1383-1400, 2010.

[41] U. A. Khan, S. Kar, and J. M. F. Moura, "Distributed sensor localization in random environments using minimal number of anchor nodes," *IEEE Trans. signal Processing*, 57: 5, pp. 2000-2016, 2009.

[42] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and qantization effects," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2506-2517, 2009.

[43] D. Acemoglu, A. Ozdaglar and A. ParandehGheibi, "Spread of (Mis)information in social networks," *Games and Economic Behavior*, vol. 70, no. 2, pp. 194-227, 2010.

[44] D. Acemoglu, G. Como, F. Fagnani, A. Ozdaglar, "Opinion fluctuations and persistent disagreement in social networks," in *IEEE Conference on Decision and Control*, pp. 2347-2352, Orlando, 2011.